

学校编码: 10384

分类号 _____ 密级 _____

学号: 23320101153131

UDC _____

厦 门 大 学

硕 士 学 位 论 文

基于单目视觉的手势检测与跟踪算法研究

Research on Hand Gesture Detection and Tracking

Based on Monocular Vision

卢 鹏

指导教师姓名: 汤碧玉 高级工程师

郑灵翔 高级工程师

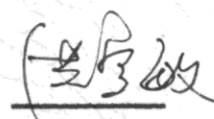
专 业 名 称: 通信与信息系统

论文提交日期: 2013 年 月

论文答辩时间: 2013 年 月

学位授予日期: 2013 年 月

答辩委员会主席:



评 阅 人: _____

2013 年 5 月

厦门大学学位论文原创性声明

本人呈交的学位论文是本人在导师指导下，独立完成的研究成果。本人在论文写作中参考其他个人或集体已经发表的研究成果，均在文中以适当方式明确标明，并符合法律规范和《厦门大学研究生学术活动规范（试行）》。

另外，该学位论文为()课题(组)的研究成果，获得()课题(组)经费或实验室的资助，在()实验室完成。(请在以上括号内填写课题或课题组负责人或实验室名称，未有此项声明内容的，可以不作特别声明。)

声明人(签名):

卢朋鸟

2013年6月3日

厦门大学博硕士论文摘要库

厦门大学学位论文著作权使用声明

本人同意厦门大学根据《中华人民共和国学位条例暂行实施办法》等规定保留和使用此学位论文，并向主管部门或其指定机构送交学位论文（包括纸质版和电子版），允许学位论文进入厦门大学图书馆及其数据库被查阅、借阅。本人同意厦门大学将学位论文加入全国博士、硕士学位论文共建单位数据库进行检索，将学位论文的标题和摘要汇编出版，采用影印、缩印或者其它方式合理复制学位论文。

本学位论文属于：

() 1. 经厦门大学保密委员会审查核定的保密学位论文；
于 年 月 日解密，解密后适用上述授权。

() 2. 不保密，适用上述授权。

（请在以上相应括号内打“√”或填上相应内容。保密学位论文应是已经厦门大学保密委员会审定过的学位论文，未经厦门大学保密委员会审定的学位论文均为公开学位论文。此声明栏不填写的，默认为公开学位论文，均适用上述授权。）

声明人（签名）：

卢朋鸟

2013 年 6 月 3 日

厦门大学博硕士论文摘要库

厦门大学博硕士论文摘要库

摘 要

随着时代的变迁和技术的进步,人类已不满足于用键盘和鼠标对计算机进行操作,因此人机交互技术就成为了当下的研究热门,而手势交互就是其中最受关注的一个方向。手是人体最为灵活的部位,其所能够表现出来的手势是复杂多变的,这对手势交互技术提出了很高的要求。手势检测和跟踪作为手势交互的前提和基础,需要克服复杂环境中的种种干扰,准确的定位出目标手势所在位置并获得一系列连续的跟踪轨迹,这是一个十分具有挑战性的课题,值得不断的深入研究。针对当前手势交互技术的发展情况,本文拟研究复杂环境下的手势检测与跟踪问题。

这里的手势检测指的是手势的第一帧自检测。为了能够在复杂环境下具有较高的鲁棒性和准确率,本文采用基于 Haar-like 特征的 AdaBoost 分类器进行手势检测,检测目标为握拳手势,检测到的若干个待定目标进一步结合时间和空间上的限制条件进行辅助判断,最后得到一个满足用户交互意愿的跟踪手势。实验结果表明该方法能够在背景中存在手势干扰的情况下准确的锁定目标手势。

手势跟踪部分的跟踪对象同样是固定手势,采用 TLD 框架进行跟踪。TLD 算法其跟踪鲁棒性高,又有 PN 在线学习方法进一步提高跟踪准确率,但是运算量过大导致实时性不好。针对这一点,本文设计了一种实时性更高的算法 BP-TLD,通过对 TLD 中数据量最大的检测模块添加反向投影算法,提高了约 20% 的跟踪速度。此外还可以根据手势跟踪的应用场景按需优化 TLD 的内部参数。从 BP-TLD 与其他手势跟踪算法 HandVu 和 MSEPF 的对比实验中可以看出,在复杂环境和高速手势的情况下 BP-TLD 的跟踪成功率要远高于其他两个算法。

在上述手势检测和跟踪算法的基础上,本文构建了一套手势控制视频播放的应用案例,演示结果表明我们设计的算法检测准确率高,跟踪稳定,不易受到复杂背景的干扰,且跟踪速度能够达到 20 帧/秒,可以应用在类触屏手势操作方面,比如视频控制、模拟鼠标和体感游戏等。

关键词: 人机交互; 手势检测; 手势跟踪

厦门大学博硕士论文摘要库

Abstract

With the times changing and technology advancing, humans are no longer satisfied with the keyboard and mouse to operate computer. Human Computer Interaction technology has become the current hot research project and gesture interaction is one of the most attractive directions. The hand is the most flexible part of the human organs, thus the gesture demonstrated by hand is complex, so the gesture interaction technology is highly on demand.

As the premise and basis of the gesture interaction, gesture detection and tracking need to overcome all the interference in a complex environment, accurately determine the location of the target gestures and a series of continuous trajectory tracking. This is a very challenging task and worthy of in-depth studies. Against the development of technologies of gesture detection and tracking, a set high robustness and real-time systems working in complex environmental interference was be promoted in this paper.

Hand gesture detecting is self-test of the first frame about the gesture. The key is to set detecting target to be a fist. Certain determined gestures are detected by AdaBoost classifier based on Haar-like feature. Combining it with the restrictions of time and space environment, we finally get an accurate tracking gesture. Experimental results show that the gesture detection algorithm, with the presence of case of gesture interference on the background, accurately target gestures. On the skilled operation lock time can be controlled within about 2 seconds.

Gesture tracking object is also a fist gesture based on TLD framework for tracking. Although TLD tracking algorithm has a high robustness and the PN online learning method further improve the tracking accuracy rate, excessive computing results no real-time. Against above weakness, this paper proposed a real-time algorithm BP-TLD, by adding back projection algorithm to the most amount of TLD detection module, to improve the tracking speed about 20%. In addition, according to the

gesture tracking application scenarios it can optimize the internal parameters of the TLD on demand. From the comparative experiments of BP-TLD and other gestures tracking algorithm HandVu and MSEPF, BP-TLD tracking the success rate is higher than the other two algorithms in the complex and high-speed gestures environment.

Based on gesture detection and tracking algorithm above, this paper constructs a set of application case of video control system, the demo results show that our proposed algorithm has highly detecting accuracy, stable tracking, less susceptible to interferences from complex background, and the tracking speed to up to 20 frames per second. So it can be used in application like touch screen gestures.

Key words: human computer interaction; hand gesture detection; hand gesture tracking

目 录

摘 要.....	I
Abstract.....	III
第 1 章 绪论.....	1
1.1. 课题研究的背景与意义.....	1
1.2. 手势检测与跟踪技术研究综述.....	2
1.2.1. 手势检测.....	4
1.2.2. 手势跟踪.....	6
1.3. 本文的研究内容及结构安排.....	10
1.3.1. 研究内容.....	10
1.3.2. 结构安排.....	10
第 2 章 手势检测算法.....	13
2.1. 概述.....	13
2.2. Haar-like 特征.....	14
2.3. 积分图.....	14
2.4. AdaBoost 分类器.....	15
2.4.1. 强分类器的训练过程.....	16
2.4.2. 级联分类器的训练过程.....	18
2.5. 改进的手势自检测算法.....	20
2.6. 实验结果与分析.....	22
2.7. 本章小结.....	24
第 3 章 手势跟踪算法.....	25
3.1. 概述.....	25
3.2. TLD 算法.....	25
3.2.1. 目标模型.....	26

3.2.2. 检测模块.....	27
3.2.3. 跟踪模块.....	30
3.2.4. 综合模块.....	30
3.2.5. 学习模块.....	31
3.3. 反向投影.....	33
3.4. BP-TLD 算法.....	35
3.5. 实验结果与分析.....	38
3.5.1. 纵向对比.....	38
3.5.2. 横向对比.....	41
3.6. 本章小结.....	43
第 4 章 视频播放控制系统.....	45
4.1. 概述.....	45
4.2. 手势交互系统设计.....	45
4.3. 本章小结.....	50
第 5 章 总结与展望.....	51
5.1. 本文工作总结.....	51
5.2. 后续工作展望.....	52
参 考 文 献.....	53
致 谢.....	61
攻读硕士学位期间发表的论文及所做工作.....	63

Contents

Abstract in Chinese	I
Abstract in English	III
Chapter 1 Preface	1
1.1. Research Background and Significance	1
1.2. Hand Gesture Detection and Tracking Methods Review	2
1.2.1. Hand Gesture Detection	4
1.2.2. Hand Gesture Tracking	6
1.3. Main Contents and Structure of this Thesis	10
1.3.1. Main Contents of This Thesis	10
1.3.2. Structure of This Thesis	10
Chapter 2 Hand Gesture Detection Method	13
2.1. Introduction	13
2.2. Haar-like Feature	14
2.3. Integral Image	14
2.4. AdaBoost Classifier	15
2.4.1. Learning Classification Functions	16
2.4.2. The Attentional Cascade	18
2.5. The Improved Hand Detection Algorithm	20
2.6. Experimental Result and Analysis	22
2.7. Chapter Summary	24
Chapter 3 Hand Gesture Tracking Method	25
3.1. Introduction	25
3.2. Tracking-Learning-Detection	25
3.2.1. Object Model	26

3.2.2. Object Detector	27
3.2.3. Object Tracker.....	30
3.2.4. Integrator	30
3.2.5. Learning Component.....	31
3.3. Back Projection.....	33
3.4. BP-TLD Algorithm.....	35
3.5. Experimental Result and Analysis	38
3.5.1. Vertical Comparison.....	38
3.5.2. Horizontals Comparison	41
3.6. Chapter Summary	43
Chapter 4 Video Control System.....	45
4.1. Introduction	45
4.2. The Design of Hand Gesture Interaction System	45
4.3. Chapter Summary	50
Chapter 5 Conclusion and Prospect.....	51
5.1. Conclusion.....	51
5.2. Prospect.....	52
Reference.....	53
Acknowledgement.....	61
Published Paper and Research during Pursuing Master Degree	63

第1章 绪论

1.1. 课题研究的背景与意义

跨入 21 世纪后, 遵循着摩尔定律的计算机不仅性能越来越强大, 而且价格也越来越便宜, 由此计算机得到了大规模的普及。进而计算机开始在社会中扮演着越来越重要的角色, 由此人们开始思考如何让计算机能够更加适应人类的生活, 特别是在人机交互 (Human-Machine Interaction, HMI) 这一方面。长久以来, 人们习惯于通过鼠标键盘的方式进行输入, 但这毕竟不符合人类自然交互的方式, 于是触摸感应、体感操作、语音识别甚至眼球、脑电波控制等新的交互方式便应运而生。比如 iPhone 就是一个鲜明的例子, 从某种程度上讲 iPhone 就是现代计算机的一个缩影, 它拥有完整的计算机架构, 但却创新性的抛弃了实体键盘, 代之以更优异的多点触摸感应技术, 一经推出立即引发了手机行业新一轮地震, 而它本身也成为了人机交互方式的一个典范。再如微软于 2010 年发布的 Kinect 体感外设, 在销售前 60 天内, 就卖出了八百万部^[1], 可见其在水机交互上的革命性进步对消费者有着巨大的吸引力。这一切活生生的事例正像卡内基·梅隆大学的 Dan R.Olsen 教授说过的: “人机交互是未来的计算机科学。我们已经花费了至少 50 年的时间来学习如何制造计算机以及如何编写计算机程序, 下一个新领域自然是让计算机服务并适应于人类的需要, 而不是强迫人类去适应计算机。”^[2]这是一个交互变革的年代, 谁提升了用户的交互感受谁就抓住了先机。

手势交互则是人机交互中的一个重要组成部分, 利用手势进行交互具有许多优点: 首先手势交互对于世界不同人之间具有很好的通用性; 其次在一些特殊的应用场景中手势交互能够得到灵活的运用; 然后手还是人类肢体表现力最强的部位, 能够表达的信息量是巨大的; 最后手势交互体验对于人类来说是自然而舒适的, 在连续的手势使用后不容易疲劳。手势按其输入设备的不同可划分为基于数据手套的手势和基于计算机视觉的手势。前者通过在手部佩戴第三方辅助设备使得计算机能够方便的识别并追踪手势信息, 在可靠性与精确度上有着天然的优势, 但是辅助设备对于人类手部的束缚和其昂贵的价格是阻碍这一技术推广的重要

原因。与之相反，基于视觉的手势识别技术只需要普通摄像头和相应的配套软件即可，成本十分低廉，不足之处就在于跟踪识别效果上有待提高。如果该类方法在技术上能够有所突破的话，未来只需通过软件上的升级即可大规模推广，其前景是不可限量的。基于计算机视觉的手势从交互的角度来讲又可分为静态手势与动态手势，静态手势通过瞬时的手的结构状态来表达某种概念层次的语义信息，动态手势则是依靠手运动的轨迹，是以随时间变化的空间特征来表述的。本文的主要研究内容是在单目视觉下的单手势检测与跟踪。

手势检测与跟踪是手势交互系统的前提和基础，如果手势检测出现了错误，后续的手势跟踪和识别就都失去了意义，而如果手势跟踪不稳定，识别出来的结果就会是曲解的。虽然人们已经意识到手势检测与跟踪的重要性，也提出了许多方法，但是在目前这些方法均存在着各种各样的问题，比如对图像背景环境的复杂程度有所限制、光照要求稳定且不能太强也不能太弱、手势运动速度不宜过快、容易受到类肤色物体的干扰等。这一系列的问题都影响到了手势交互系统的实用性，因此本文对复杂背景下的手势检测和跟踪算法进行研究具有重要的理论意义和实用价值。

1.2. 手势检测与跟踪技术研究综述

手势交互技术按流程可简略的划分为手势检测、手势跟踪、手势识别和多媒体交互四个部分，如图 1-1 所示。手势检测模块通过判断用户的交互意愿来锁定第一帧手势。在第一帧手势信息的基础上，跟踪模块从连续的图像序列提取手势轨迹。接着手势识别模块对获得的静态手势或动态轨迹进行解释，使其与预设的表达意思一一对应，最后得到的表达意思以控制命令的形式输入到多媒体交互系统中，并实现相应的多媒体效果反馈给用户。本文的研究内容主要是手势检测和手势跟踪两方面。

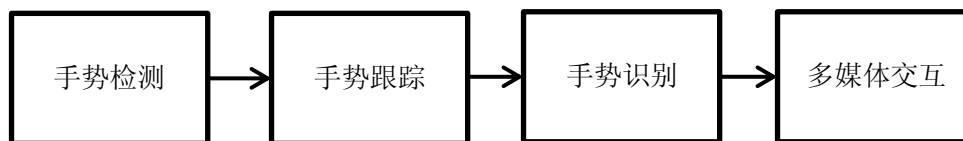


图 1-1 手势交互技术

Degree papers are in the "[Xiamen University Electronic Theses and Dissertations Database](#)". Full texts are available in the following ways:

1. If your library is a CALIS member libraries, please log on <http://etd.calis.edu.cn/> and submit requests online, or consult the interlibrary loan department in your library.
2. For users of non-CALIS member libraries, please mail to etd@xmu.edu.cn for delivery details.

厦门大学博硕士论文摘要库